GROUP-TESTING WITH AT MOST D DEFECTIVES: BINOMIAL AND HYPERGEOMETRIC MODELS*

by

Milton Sobel

Technical Report No. 204

April, 1973

University of Minnesota

Minneapolis, Minnesota

* Supported by U.S. Army Grant DA-ARO-D-31-124-72-6187 at the University of Minnesota, Minneapolis, Minnesota.

1. INTRODUCTION

We consider a group-testing problem in which each of N units has to be classified into one of two disjoint, exhaustive states, which we call satisfactory and defective. Any integer number x of units can be tested simultaneously with only two possible responses:

i) all x are satisfactory.

ii) at least one of the x is defective.

In the latter case we don't know at this point how many or which ones are defective.

In this paper we assume that it is known a priori or given to us that the N units contain at most D defectives; the problem is to find an efficient way of using that information and to study the effect of this information on the number of tests required.

Two models are considered and in both we use $n \leq N$ to denote the current number of units (among the N) that are not yet classified and $d \leq D$ to denote the current upper bound on the number of defectives among the n units. Model 1 is the conditional binomial model and Model 2 is the conditional hypergeometric model; in both cases the condition is that the number of defectives among the n units is at most d. In Model 1 we let $q > 0$ (resp., $p = 1 - q$) denote the known, unconditional probability that a unit is satisfactory (resp., defective). In Model 2 we introduce into the model a finite population of known size $K \geq N + D$ that contains exactly D defectives. Then the sample of size N taken from the K by random sampling contains at most D defectives and need not contain any defectives. The current sample of size n can then be regarded as a random sample from a finite population of size $k = K - (N-n)$, it contains at most d defectives and need not contain any defectives. For the current values we note that $k - n = K - N \geq D \geq d$, so that $0 \leq n \leq k - d \leq k$.

As in other group-testing problems (of [3], [4] and [5]) we look for an optimal solution in a class of group-testing procedures that we call "nested, with recombination", NWR. "Nested" means that we always prefer to break up sets that are known to contain at least one defective and "Recombination" means that if two sets can be combined before the next test without any loss of information (the probability distribution after combining is the same as before), then we do this and call it recombination.

The optimal NWR procedure is found for both Model 1 and Model 2; we show that, under a simple linear correspondence between the parameter p in Model 1 and the parameter K in Model 2, the two optimal strategies, as well as the resulting expected number of tests, are identical. Various results and tables are given for this procedure, which we call R_{1D} (resp., R_{2D}) as it applies to Model 1 (resp., Model 2). For some fixed values of q in Model 1 we give numerical results up to $N = 50$. Lower bounds for the expected number of tests in this problem are also derived for both models in Section 4.

Some of the motivation for Model 1 of this paper came from a recent paper of Thomas, Pasternak, Vacirca, and Thompson [6] in which group-testing methods are applied to a hospital problem of finding "leaking" vials in a drawer containing 50 vials of radioactive materials. Here without any statistical formulation and with only empirical justification it is assumed that each set of $N = 50$ vials can contain at most one defective (i.e., leaking) vial. The implications of this assumption are not studied in [6] and, as a result, unfair comparisons are made there between a procedure that classifies all units correctly without uncertainty and another (their procedure R_T) that classifies all units correctly only if the assumption holds that there is at most $D = 1$ defective in each set of $N = 50$ vials.

As an application of Model 1 above, we try to put the method in [6] on a sound basis by providing a formulation under which the optimal NWR procedure reduces to procedure R_T in some cases, (but not in all cases, as is illustrated in Section 5). Procedure R_T (defined only for $D = 1$ in [6]) is simply to start by testing all the N units and if it is positive (i.e., if the test shows at least one defective), then we use the so-called "halving procedure" that is known (cf. [2] and Appendix C of [3]) to be optimal for finding the defective, when we know that there is exactly one defective present. The efficiency of this procedure R_T relative to the optimal NWR procedure $R_{1,1}$ for the case of at most $D = 1$ is investigated in Section 5 .

The (unfair) comparisons mentioned above lead to results for the expected number of tests under procedure R_T that are better, i.e., lower, than the lower bound for any group-testing procedure that (unconditionally) classifies all units correctly. It is therefore very desirable to derive valid lower bounds that hold for procedure R_T and also for any procedure that uses the assumption of an upper bound D on the number of defectives present among the N units. An additional purpose of this is to evaluate the efficiency of R_T and of $R_{1,D}$ among all the procedures that use the knowledge of the upper bound D.

2. PRELIMINARIES FOR MODELS 1 AND 2

Under Model 1, we use n and d for the current values of N and D, respectively, and let (noting that $d \le n$ but D can be $> n$)

$$(2.1)\qquad B_\delta(n) = \sum_{\alpha=0}^{\mathrm{Min}(\delta,n)} \binom{n}{\alpha} p^\alpha q^{n-\alpha} \qquad\qquad (\delta = d, D).$$

The probability that a random sample of size x taken from n has no defectives is (for the H-situation)

$$(2.2)\qquad P_{1H} = P\{X = 0 | n,d\} = \frac{q^x B_d(n-x)}{B_d(n)},$$

where we note from (2.1) that for $x = n$ the value of $B_d(0) = 1$ and we are using X to denote the number of defectives in the sample of size x.

If a set of size m ($\le n$) already has at least one defective and we take a random sample of size x ($< m$) from it, then the probability P_{1G} that it is free of defectives needs to be derived. Let b (resp., β) denote the exact number of defectives among all n unclassified units (resp., in the set of size m). The denominator of our desired conditional probability P_{1G} is the probability that the so-called defective set of size m has at least one defective (note that $\beta \geq 1$ below) and this is

$$(2.3)\qquad \sum_{\gamma=0}^{d}\left[\sum_{\beta=1}^{\gamma}\binom{m}{\beta}\binom{n-m}{\gamma-\beta}\right]p^\gamma q^{n-\gamma} = \sum_{\gamma=0}^{d} p^\gamma q^{n-\gamma}\left[\binom{n}{\gamma} - \binom{n-m}{\gamma}\right] = B_d(n) - q^m B_d(n-m)$$

by the use of the well-known hypergeometric identity (which is also used below). The desired conditional probability P_{1G} is then given by

$$(2.4)\qquad P_{1G} = \sum_{b=0}^{d} \frac{\sum_{\beta=1}^{b}\binom{m-x}{\beta}\binom{n-m}{b-\beta}p^b q^{n-b}}{B_d(n) - q^m B_d(n-m)} = \frac{q^x B_d(n-x) - q^m B_d(n-m)}{B_d(n) - q^m B_d(n-m)}.$$

For Model 2 it is well known that taking a random sample of size $x \le N$ from the N units, which is in turn a random sample from a larger set of size K, is equivalent to taking a random sample of size x from the larger set of size K; the fact that the latter set has exactly D defectives does not affect this result. We also use this result for the current situation with (N,D,K) replaced by (n,d,k), respectively, where $k = K - (N-n) \ge n$.

The probability P_{2H} that a random sample of size x (taken from k units which contain exactly d defectives) is free of defectives is given by

$$(2.5) \qquad P_{2H} = \frac{\binom{k-x}{d}}{\binom{k}{d}} = \frac{\binom{k-d}{x}}{\binom{k}{x}},$$

which depends on K, N and n only through k. We have seen in Section 1 that $n \le k - d$ and hence $x \le n \le k - d$, so that P_{2H} in (2.5) is never zero.

The denominator of the desired conditional probability P_{2G} for the G-situation is the probability that the subset of size m (taken as a random sample from the set of size k) has at least one defective unit (note that $\alpha \ge 1$ below) and this is given by

$$(2.6) \qquad \sum_{\alpha=1}^{d} \frac{\binom{d}{\alpha}\binom{k-d}{m-\alpha}}{\binom{k}{m}} = \frac{\binom{k}{m} - \binom{k-d}{m}}{\binom{k}{m}} = D_{2G} \quad \text{(say)},$$

which is one for $m > k - d$. However, since $n \le k - d$, it follows that $m \le n \le k - d$ and hence we never encounter such m-values. Then the desired conditional probability P_{2G} that a random sample of size x (taken from the so-called defective set of size m) is free of defectives is given by

$$(2.7) \qquad P_{2G} = \frac{1}{D_{2G}} \sum_{\alpha=1}^{d} \frac{\binom{d}{\alpha}\binom{k-d}{m-\alpha}}{\binom{k}{m}} \cdot \frac{\binom{m-\alpha}{x}}{\binom{m}{x}} = \frac{1}{D_{2G}} \sum_{\alpha=1}^{d} \frac{\binom{k-m}{d-\alpha}\binom{m-x}{\alpha}}{\binom{k}{d}}$$

$$= \frac{1}{D_{2G}} \left\{ \frac{\binom{k-x}{d} - \binom{k-m}{d}}{\binom{k}{d}} \right\}$$

Since D_{2G} from (2.6) can be written in two ways as

$$(2.8) \qquad D_{2G} = 1 - \frac{\binom{k-d}{m}}{\binom{k}{m}} = 1 - \frac{\binom{k-m}{d}}{\binom{k}{d}} = \frac{\binom{k}{d} - \binom{k-m}{d}}{\binom{k}{d}},$$

it follows that we can also write P_{2G} in the form

$$(2.9) \qquad P_{2G} = \frac{\binom{k-x}{d} - \binom{k-m}{d}}{\binom{k}{d} - \binom{k-m}{d}} = 1 - \frac{\binom{k}{d} - \binom{k-x}{d}}{\binom{k}{d} - \binom{k-m}{d}}$$

These expressions (P_{1H}, P_{1G}) and (P_{2H}, P_{2G}) are essential ingredients for derivin[g] the basic algorithm that implicitly defines our proposed NWR procedure for Model 1 and Model 2, respectively in the next section.

3. THE ALGORITHM: ORIGINAL AND STANDARD FORM

For Model 1 let $H_d(n) = H_d(n;q)$ denote the expected number of additional tests needed under procedure R_{1D} when there are n unclassified units remaining with at most d defectives and we have no further information (except, of course, that q is known). Let $G_d(m,n) = G_d(m,n;q)$ denote the same, except that a particular subset of size $m \le n$ is known to contain at least one defective; for $m = 1$ we have boundary conditions below and for $m = 0$ the $G_d(m,n)$-function coincides with the $H_d(n)$-function. In line with group-testing terminology we call the case $m = 0$ an H-situation and the case $m \ge 2$ is called a G-situation. Under Model 1 for $n \ge 1$ and $d \ge 1$ in the H-situation

$$(3.1) \qquad H_d(n) = 1 + \min_{1 \le x \le n} \left\{ \frac{q^x B_d(n-x)}{B_d(n)} H_d(n-x) + \left[\frac{B_d(n) - q^x B_d(n-x)}{B_d(n)} \right] G_d(x,n) \right\} .$$

Under Model 1 for $n \ge m \ge 2$, $d \ge 2$ and also for $n = m \ge 2$, $d = 1$ in the G-situation (the case $n > m, d = 1$ is a boundary condition below)

$$(3.2) \qquad G_d(m,n) = 1 + \min_{1 \le x < m} \left\{ \left[\frac{q^x B_d(n-x) - q^m B_d(n-m)}{B_d(n) - q^m B_d(n-m)} \right] G_d(m-x,\ n-x) + \left[\frac{B_d(n) - q^x B_d(n-x)}{B_d(n) - q^m B_d(n-m)} \right] G_d(x,\ n) \right\}.$$

The boundary conditions for Model 1 are threefold:

$$(3.3) \qquad H_d(n) = 0 \qquad \text{if} \quad d = 0 \quad \text{or} \quad n = 0,$$

$$(3.4) \qquad G_d(1,\ n) = H_{d-1}(n-1) \qquad n = 1, 2, \ldots, \quad d \ge 1,$$

$$(3.5) \qquad G_1(m,\ n) = G_1(m,\ m) .$$

Numerical values of $H_D(N)$ are given in Table 3 at the end of this paper for $D = 1, 2, 3$, $N = 1(1)50$ and $q = .75, .90, .95,$ and $.99$.

Remark:

It might be of interest to define new functions $V_d(n)$ and $U_d(m, n)$ (in analogy with $H_d(n)$ and $G_d(m, n)$) to keep track of the number of units classified by inference. Then we remove the 1 and the Min in both (3.1) and (3.2), put $V_d(n)$ for $H_d(n)$ and $U_d(m, n)$ for $G_d(m, n)$ and use the same x-values as prescribed by procedure R_{1D} in (3.1) and (3.2), respectively. For (3.3) we write $V_d(n) = 0$ if $d = 0$ or $n = 0$. For (3.4) we write

$$(3.6) \qquad U_d(1,n) = V_{d-1}(n-1) \qquad n = 1,2,\ldots, \; d \geq 1$$

and for (3.5) we write, for $d = 1$ only,

$$(3.7) \qquad U_1(m, n) = (n-m) + U_1(m, m).$$

Then $V_D(N)$ is the required expected number of units classified by inference using (3.5) if we start with N units and at most D defectives. This technique could clearly be applied to either Model and can easily be modified so that all the units classified by inference in both (3.4) and (3.5) are counted. These calculations have not been carried out in this paper.

For Model 2 we use the symbols $H_d(n, k)$ and $G_d(m, n, k)$ to distinguish it from Model 1 and use the fact that $n \leq k - d$. Under Model 2 for $n \geq 1$ and $d \geq 1$ in the H-situation

$$(3.8) \qquad H_d(n, k) = 1 + \min_{1 \leq x \leq n} \left\{ \frac{\binom{k-x}{d}}{\binom{k}{d}} H_d(n-x, k-x) + \left[1 - \frac{\binom{k-x}{d}}{\binom{k}{d}} \right] G_d(x, n, k) \right\}.$$

Under Model 2 for $n \geq m \geq 2$, $d \geq 1$ and also for $n = m \geq 2$, $d = 1$ in the G-situation

$$(3.9) \qquad G_d(m, n, k) = 1 + \min_{1 \le x < m} \left\{ \left[\frac{\binom{k-x}{d} - \binom{k-m}{d}}{\binom{k}{d} - \binom{k-m}{d}} \right] G_d(m-x, n-x, k-x) + \left[\frac{\binom{k}{d} - \binom{k-x}{d}}{\binom{k}{d} - \binom{k-m}{d}} \right] G_d(x, n, k) \right\} .$$

The boundary conditions for Model 2 are

$$(3.10) \qquad H_d(n, k) = 0 \qquad \text{if } d = 0 \text{ or } n = 0,$$

$$(3.11) \qquad G_d(1, n, k) = H_{d-1}(n-1, k-1) \qquad n = 1, 2, \ldots, \quad d \ge 1,$$

and, for $d = 1$ only,

$$(3.12) \qquad G_1(m, n, k) = G_1(m, m, k-n+m).$$

Remark:

For the special case $d = 1$ we note several interesting things. Using (2.1) to evaluate the square brackets in (3.2), we find that (3.2) and (3.9) are identical and do not depend on q or k. Since we have at most 1 defective and at least 1 defective, we must have exactly 1 defective and it lies in the set of size m. The common algorithm of (3.2) and (3.9) has the same boundary conditions since $H_0(n-1, k-1) = 0$ in (3.11) and $G_1(m, m, k-n+m)$ in (3.12) does not depend on the last argument. In fact this algorithm leads to the well-known "halving procedure" that is known to be optimal when we are given that there is exactly 1 defective present. An exact formula for $G_1(m, m, k) = G_1(m)$ given in e.g. (2.22) of [5] is

$$(3.13) \qquad G_1(m) = b + 2 - \frac{2^{b+1}}{m} ,$$

where b is the integer part of $\log_2 m$. If we also compare (3.1) and (3.8) for $d = 1$ we find that q and k do not cancel out but the two algorithms are identical if we identify $k = K - N + n$ in Model 2 with $n - 1 + (1/p)$ in Model 1, i.e., if we identify p with $(K-N+1)^{-1}$. Since the latter does not depend on n, it means that one simple substitution will relate the entire procedure in one model with the entire procedure in the other model. It also follows that the optimal NWR strategy or set of x-values for any $D \geq 1$ and $d = 1$ and for all situations is the same under both models; indeed, this is the main reason for including both models in the same paper.

In general, for values of $D > 1$ the two optimal NWR procedures under Models 1 and 2 cannot be related in exactly the same way. If $N \leq \mathrm{Min}(D, K-D) = K'$ (say) then this "information" is redundant under Model 1 and the resulting NWR procedure does not even depend on D (or d). Under Model 2 the optimal NWR procedure keeps track of d (which equals the known D at the outset) even when $N \leq K'$ and uses both d and n to determine what x-value is optimal. It appears to be the case that for any $N \leq K'$ the two optimal nested procedures are related as follows. If we take optimal trees and the polynomial (in q) expressions for $H_1(N) = H_1(N; q)$ for procedure R_1 from [3] for the binomial problem (without any D) and replace p^j by $D^{[j]}/K^{[j]}$ $(j=0, 1, 2, \ldots)$ where $D^{[j]} = D(D-1)\ldots(D-j+1)$, then the same tree and the resulting expressions also hold for the hypergeometric model; new dividing points have to be obtained however. For example if $D \geq 2$ and $N = 2$ then the result from [3] for the binomial model is

$$(3.14) \qquad H_1(2) = \begin{cases} 1 + 3p - p^2 & \text{for } 0 \leq p \leq \frac{3-\sqrt{5}}{2} = p_0 \text{ (say); take } x = 2 \\ 2 & \text{for } p_0 \leq p \leq 1; \qquad\qquad \text{take } x = 1. \end{cases}$$

The corresponding result for the hypergeometric model with $2 \le D \le K-2$, $K \ge 6$ is

$$(3.15) \qquad H_{2,D}(2) = \begin{cases} 1 + 3\frac{D}{K} - \frac{D(D-1)}{K(K-1)} & \text{for } 2 \le D \le \frac{3K-2-\sqrt{5K^2-8K+4}}{2}; \text{ take } x = 2 \\ 2 & \text{for } K-2 \ge D \ge \frac{3K-2-\sqrt{5K^2-8K+4}}{2}; \text{ take } x = 1. \end{cases}$$

If $K = 4$ or 5 then we omit the first line in (3.15) and take $x = 1$. The dividing point shown above is obtained by simply equating the two expressions above. In short, the strategy or tree to be used is the same but the resulting expected number of tests and dividing points are expressed in terms of p (or q) in one case and in terms of K and D in the other.

It is interesting to note that the same technique that was used above with p can also be used with q by replacing q^j by $(K-D)^{[j]}/K^{[j]}$ ($j = 0, 1, 2\ldots$). These two techniques give consistent results because of the interesting identity for any integer $m < A + 1$ (assume $A \ge B$, for convenience)

$$(3.16) \qquad \frac{(A-B)^{[m]}}{A^{[m]}} = \sum_{\alpha=0}^{m} (-1)^{\alpha} \binom{m}{\alpha} \frac{B^{[\alpha]}}{A^{[\alpha]}}$$

where $B^{[0]} \equiv 1$. This follows from the known identity

$$(3.17) \qquad (A-B)^{[m]} = \sum_{\alpha=0}^{m} (-1)^{\alpha} \binom{m}{\alpha} (A-\alpha)^{[m-\alpha]} B^{[\alpha]}$$

whose proof is easy by induction (and is omitted), and the fact that $A^{[m]} = A^{[\alpha]}(A-\alpha)^{[m-\alpha]}$ for every α ($\alpha = 0,1,\ldots,m$).

For the standard form under Model 1 with any $d \ge 1$ we define for $n \ge 1$, $B_d^*(n) = B_d(n)/q^{n-1}$ (Note that $B_d^*(0) = q > 0$ for all $d \ge 0$),

$$(3.18) \qquad H_d^*(n) = B_d^*(n)\, H_d(n),$$

$$(3.19) \qquad G_d^*(m, n) = [B_d^*(n) - B_d^*(n-m)]\; G_d(m, n).$$

In terms of these new quantities the algorithm takes the simpler form:

for $n \geq 1$

(3.20) $$H_d^*(n) = B_d^*(n) + \operatorname*{Min}_{1\leq x\leq n} \{ H_d^*(n-x) + G_d^*(x, n) \};$$

for $n \geq m \geq 2$, $d > 1$ and also for $n = m \geq 2$, $d = 1$

(3.21) $$G_d^*(m, n) = B_d^*(n) - B_d^*(n-m) + \operatorname*{Min}_{1\leq x<m} \{ G_d^*(m-x, n-x) + G_d^*(x, n) \};$$

the boundary conditions are

(3.22) $$H_d^*(n) = 0 \qquad \text{if } d = 0 \text{ or } n = 0,$$

(3.23) $$G_d^*(1, n) = [B_d^*(n) - B_d^*(n-1)]\, G_d(1, n) = \frac{p}{q} B_{d-1}^*(n-1) = \frac{p}{q} H_{d-1}^*(n-1) (n\geq 1, d\geq 1),$$

and, for $d = 1$ only,

(3.24) $$G_1^*(m,n) = [B_1^*(n) - B_1^*(n-m)]\, G_1(m,n) = \left[\frac{B_1^*(n) - B_1^*(n - m)}{B_1^*(m) - q}\right] G_1^*(m,m) = G_1^*(m,m).$$

It is interesting to note that for $d = 1$ the value of $B_1^*(n) = 1 + (n-1)p$ is linear in p and the right side of (3.23) is zero by (3.22). Hence the linearity in p is preserved as we increase n, keeping $d = 1$. Results for $d = 1$ are given at the end of this paper. Some conjectures about the optimal strategy for $d = 1$ will be given in a later section.

For the standard form under Model 2 we define

(3.25) $$H_d^*(n, k) = \binom{k}{d} H_d(n, k)$$

(3.26) $$G_d^*(m, n, k) = \left[\binom{k}{d} - \binom{k-m}{d}\right] G_d(m, n, k)$$

and we recall that $m \leq n \leq k - d$ since $N \leq K - D$.

The algorithm takes the simpler form: for $n \geq 1$

(3.27) $$H_d^*(n, k) = \binom{k}{d} + \min_{1 \leq x \leq n} \{ H_d^*(n-x, k-x) + G_d^*(x, n, k) \};$$

for $n \geq m \geq 2$, $d > 1$ and also for $n = m \geq 2$, $d = 1$

(3.28) $$G_d^*(m,n,k) = \binom{k}{d} - \binom{k-m}{d} + \min_{1 \leq x < m} \{ G_d^*(m-x, n-x, k-x) + G_d^*(x,n,k) \};$$

the boundary conditions are

(3.29) $$H_d^*(n, k) = 0 \qquad \text{if } d = 0 \text{ or } n = 0,$$

(3.30) $$G_d^*(1,n,k) = \binom{k-1}{d-1} G_d(1,n,k) = \binom{k-1}{d-1} H_{d-1}(n-1,k-1) = H_{d-1}^*(n-1,k-1) \quad (n \geq 1,\ d \geq 1),$$

and, for $d = 1$ only,

(3.31) $$G_1^*(m,n,k) = mG_1(m,n,k) = mG_1(m,m,k-n+m) = G_1^*(m,m,k-n+m) = G_1^*(m,m,m).$$

The last result in (3.31) follows by inference also since we have at most 1 defective left and at least one among the m units. In other words $G_1^*(m,n,k) = G_1^*(m)$ depends only on the first argument and indeed for $d = 1$ this leads to the halving procedure; the value of $G_1^*(m)$ is m times that for $G_1(m)$ given in (3.13) and is a basic integer in group-testing.

Remark:

It should be pointed out that the assumption that $N \leq K - D$ (which implies that $n \leq k - d$) was made for Model 2 above only for convenience. Without this assumption we have to add to the boundary conditions the fact that whenever $d = k \geq 1$ we can infer that $n = d = k$ and the remaining n units are all defective, i.e., we add the condition

(3.32) $$H_k^*(k,k) = H_k(k,k) = G_k(m,k,k) = G_k^*(m,k,k) = 0 \qquad \text{for all } m.$$

In addition, we will not take any x for which the outcome of a test is known in advance; hence $x \leq k - d$ in both situations. In short, if we make all possible inferences as we go along, then the above algorithm can also be applied when $N < K - D$ in Model 2.

When $N \leq D$ then the procedure $R_{1,D}$ is identical with the procedure R_1 in [3] and [4]. When $N \leq \mathrm{Min}(D, K - D)$ then the procedure $R_{2,D}$ is equivalent to procedure R_1 in the sense of a transformation between K and p. As N increases above this value and approaches K the procedure $R_{2,D}$ becomes more and more similar to the procedure $R_1(S, D)$ studied in [5], where we know exactly (at the outset) the number of defectives D and the number of satisfactory units $S = N - D$ among the N units, but still have to find out which ones are the defective units.

4. LOWER AND UPPER BOUNDS ON THE EXPECTED NUMBER OF TESTS

For lower bounds we use the two usual methods [4]. One, based on information theory, gives us the information lower bound ILB and the other, based on coding theory, is called the Huffman lower bound, HLB. Although the latter is always better, i.e., larger, it has no explicit analytic formula like the ILB and, since the HLB is based on a finite N (and is not linear in N), it fails to yield a useful result for the case $N = \infty$. Hence we have an interest in both of these lower bounds.

Under Model 1 we use the fact that

$$(4.1) \qquad \sum_{\alpha=0}^{d} \binom{n}{\alpha} p^{\alpha} q^{n-\alpha} = I_q(n-d,\ d+1)\ ,$$

where $I_q(a,b)$ denotes the usual incomplete beta function and $I_q(0,b) = 1$ for any $b > 0$. Since there $\binom{n}{\alpha}$ states of nature each with probability $p^{\alpha} q^{n-\alpha}/I_q(n-d,\ d+1)$ $(\alpha = 0, 1, \ldots, d)$, the total information T_1 to be taken out of the sample of size n is

$$(4.2) \qquad T_1 = -\sum_{\alpha=0}^{d} \binom{n}{\alpha} \frac{p^{\alpha} q^{n-\alpha}}{I_q(n-d,\ d+1)} \log_2 \left\{ \frac{p^{\alpha} q^{n-\alpha}}{I_q(n-d,\ d+1)} \right\}$$

$$= \log_2 I_q(n-d,\ d+1) - \frac{np \log_2 p}{I_q(n-d,\ d+1)} \sum_{\beta=0}^{d} \binom{n-1}{\beta} p^{\beta} q^{n-1-\beta}$$

$$- \frac{nq \log_2 q}{I_q(n-d,\ d+1)} \sum_{\beta=n-1-d}^{n-1} \binom{n-1}{\beta} q^{\beta} p^{n-1-\beta}$$

$$= \log_2 I_q(n-d,\ d+1) - n \left\{ \frac{p(\log_2 p)\, I_q(n-d,\ d) + q(\log_2 q)\, I_q(n-d-1, d+1)}{I_q(n-d,\ d+1)} \right\}.$$

Since we can get at most 1 unit of information per test, the above expression for T_1 is a lower bound on the number of tests required by any procedure and hence also on the expected number of tests required; this gives us the ILB.

For the special case $d = D = 1$ and $n = N$, straightforward integration (of the beta functions) and algebra in (4.2) yields the ILB, for Model 1

$$(4.3) \qquad T_1 = \log_2(q + Np) - \left\{ \frac{Np \log_2 p + q \log_2 q}{q + Np} \right\}$$

$$= p' \log_2 N - \{ p' \log_2 p' + q' \log_2 q' \},$$

where $p' = Np/(q+Np)$. Although (4.3) shows that T_1 is always less than $1 + \log_2 N$, a better result obtained by differentiating with respect to p' is

$$(4.4) \qquad T_1 \le \log_2 (N + 1);$$

this value is attained at $p = q = 1/2$.

The HLB consists of ordering the $C = \sum_{\alpha=0}^{d} \binom{n}{\alpha} = 2^n I_{\frac{1}{2}}(n-d, d+1)$ probabilities for the various states of nature. The smallest two are added and replaced by a single new number; the reduced set with $C - 1$ probabilities is reordered and the process is repeated. The sum of all the new numbers is the HLB [1]. For the overall result starting at the outset, we merely replace n and d by N and D, respectively, in both of these bounds.

Under Model 2 there are again $C = \sum_{\alpha=0}^{d} \binom{n}{\alpha}$ states of nature and for given α each has the same probability, namely

$$(4.5) \qquad P_\alpha = \frac{\binom{d}{\alpha}\binom{k-d}{n-\alpha}}{\binom{k}{n}\binom{n}{\alpha}} = \frac{\binom{k-n}{d-\alpha}}{\binom{k}{d}} \qquad (\alpha = 0, 1, \ldots, d).$$

Hence the total information T_2 is

$$(4.6) \qquad T_2 = -\sum_{\alpha=0}^{d} \binom{n}{\alpha} \frac{\binom{d}{\alpha}\binom{k-d}{n-\alpha}}{\binom{k}{n}\binom{n}{\alpha}} \log_2 \frac{\binom{k-n}{d-\alpha}}{\binom{k}{d}} = \log_2\binom{k}{d} - \sum_{\alpha=0}^{d-1} \frac{\binom{d}{\alpha}\binom{k-d}{n-\alpha}}{\binom{k}{n}} \log_2 \binom{k-n}{d-\alpha}.$$

For $d = D = 1$, $n = N$ and $k = K$ this yields the ILB

$$(4.7) \qquad T_2 = \log_2 K - \left(\frac{K-N}{K}\right) \log_2 (K-N);$$

$\log_2 K$ is an upper bound for T_2. It is interesting to note that if we identify p with $(K-N+1)^{-1}$ in (4.3) or set $K = N-1+(1/p)$ in (4.7) then the two ILB results in (4.3) and (4.7) become identical. A similar result should also hold for the HLB's.

To get an upper bound on the expected number of tests we use procedure R_T (cf. [6]) which always starts by testing all of the N units. This is an available strategy in our minimization plan and must yield an upper bound. In both cases we use (3.13). Under Model 1 with $n = N$ and $d = D = 1$

$$(4.8) \qquad H_D(N) < \frac{q^N}{q^N+Npq^{N-1}} (1) + \frac{Npq^{N-1}}{q^N+Npq^{N-1}} \left(b + 3 - \frac{2^{b+1}}{N}\right)$$

$$= 1 + \frac{Np}{q+Np}\left(b + 2 - \frac{2^{b+1}}{N}\right) \le b + 3 - \frac{2^{b+1}}{N},$$

where b is the integer part of $\log_2 N$. For Model 2 with $n = N$, $k = K$, and $d = D = 1$

$$(4.9) \qquad H_D(N,K) \le 1 + \frac{N}{K} \left(b + 2 - \frac{2^{b+1}}{N}\right) \le b + 3 - \frac{2^{b+1}}{N} .$$

Thus the common value on the right of (4.8) and (4.9) is an upper bound for $D = 1$ under both models, and if we identify K with $N - 1 + (1/p)$ the middle expressions in (4.8) and (4.9) will also be identical. Of course, the value N itself is also an upper bound for $H_D(N)$ under our procedure for all parameters and in both models.

5. GROUP-TESTING WITH CONFIDENCE LEVEL P^*

As an application of the group-testing problem with at most D defectives, we consider a problem in which we do not necessarily want to classify every unit correctly but we do want to control the (overall) probability of classifying all the N units correctly.

Suppose we want to have a probability of at least P^* ($P^* \le 1$ is preassigned) that all the N units are classified correctly; other criteria could also be considered here but we only consider one criteria in this paper. With the knowledge of q and N in Model 1, we define D^* by the relation

$$(5.1) \qquad B_{D^*-1}(N) < P^* \le B_{D^*}(N) \qquad\qquad (D^* \ge 1);$$

we use the symbol D^*_r for the unique randomized combination of D^*-1 and D^* that makes the corresponding combination of the extremes in (5.1) exactly equal to the specified P^*. This solution always exists for any P^* since for $D = N$ the value of $B_N(N) = 1$; we are not concerned about values of $P^* < q^N$. (It remains to be shown that the randomized combination of two consecutive integers will always give us the best results.)

Having defined D^*_r, our procedure now consists of carrying out this randomized experiment (e.g. by tossing a coin with the desired probabilities for heads and tails) to determine at the outset whether we will use $D^* - 1$ or D^* as an upper bound on the number of defectives. We then use the above procedure $R_{1,D}$ with $D = D^*$ (say), i.e., we operate on the assumption that there are at most D^* defectives among the N units. If $N \le D^* \le K-N$ then the procedure reduces to the binomial procedure R_1 [3] without any restrictions on D; if q is large then D^* will be small and we will often have $D^* = 1$. Thus

$D^* = 1$ is a most important case since it is the case with the greatest savings.

It is clear that the same formulation could be used with Model 2 if we replaced $B_D(N) = B_D(N; q)$ by $A_{D_o}(n; K, D)$ defined by

$$(5.2) \qquad A_{D_o}(N; K, D) = \sum_{\alpha=0}^{D_o} \frac{\binom{D}{\alpha}\binom{K-D}{N-\alpha}}{\binom{K}{N}} \qquad (D_o = 0, 1, \ldots, D)$$

and again the value of $A_D(N; K, D) = 1$; in (5.2) we use D_o as the derived upper bound for the N units and D is the number of defectives among the K units.

We can apply this type of formulation to the example treated by Thomas et. al. in [6], where it is assumed, without any justification or analysis of the implications, that $D = 1$. They used the binomial model with $N = 50$; $q = .99$ (and .95). Their procedure R_T is to start by testing all N units; this agrees with $R_{1,1}$ for q sufficiently large. For $q = .99$ the two procedures agree but for $q \leqq .97$ they do not. Even when our procedures agree, our calculations do not; evidently they use unconditional binomial probabilities and we use conditional binomial probabilities to compute the expected number of tests, namely either middle expression in (4.8) above. For $q = .99$ the common value for procedures $R_{1,1}$ and R_T is 2.9195 and the value of $H_1(50)/50$ is .0584; this corresponds to the value .065 in [6].

For $q = .95$ under procedure $R_{1,1}$ we start by testing 32 units and if it contains defectives we use the halving procedure; if not, we test all of the remaining 18 units. If these are all satisfactory we are through and, if not, we use the halving procedure for the 18 units. The value of $H_1(50)$ is 4.9565 for procedure $R_{1,1}$ and 5.1449 for procedure R_T (using the formula in (4.8) above) so that on a per unit basis the results are .0991 and .1029, respectively; the corresponding result in [6] is .126.

For $q = .9$ the procedure $R_{1,1}$ again starts with a test of 32 units; if they contain a defective, we use the halving procedure. If not, we test 10 of the remaining 18. If these contain a defective we use the halving procedure on these 10; if not, we test all of the remaining 8 units. If these contain a defective we use the halving procedure on these 8 units; if not we are through. Here we use 3 tests if the units are all satisfactory as opposed to procedure R_T where we only use one. The value of $H_1(50)$ is 5.4407 for procedure R_1 and 5.8474 for procedure R_T when $q = .9$.

For $q = .99$ we can use (4.3) or (4.7) to obtain the ILB and dividing by $N = 50$ (to put everything on a per unit basis), this gives .0563; the upper bound is again .0584. For $q = .95$ we obtain for the ILB .0988 and the upper bound is the value .1029 above. For $q = .90$ the ILB is .1081 and the upper bound is .1169. These values give us some idea of the efficiency of procedures $R_{1,1}$ and R_T by taking ratios with the ILB. The efficiencies (in %) for $R_{1,1}$ are 96.4, 99.6, and 99.2 for $q = .99$, .95 and .90, respectively while those for R_T are 96.4, 96.0, and 92.4 respectively.

In order to make the two procedures comparable above we either take $D = 1$ in both or equivalently set the preassigned level of P^* equal to the exact value of the probability attained by the use of procedure R_T with $D = 1$ in [6]. The latter value is calculated to be .9128; this corresponds to $D = 1$ exactly when $N = 50$ and $q = .99$. If we were to take $P^* = .90$ then we could do better, of course, but in practice one might like to use (and we recommend) much higher values of P^*; no such calculation was carried out in [6].

The above efficiencies with fixed N and fixed P^* (or with $D = 1$) for $q = .90$, .95, and .99 are not indicative of a limiting trend as $q \to 1$ since as soon as $q^N \geq P^*$ we can stop making any tests and get infinitely high efficiency. On the other hand if we let $N \to \infty$ and $p \to 0$ so that $Np \to \lambda$ and $q^N \sim e^{-\lambda}$ remains just below a fixed value of P^*, then we can keep $D = 1$ throughout the limiting process and use the ratio of (4.3) to either middle expression in (4.8) to study the limiting efficiency. Since both of these increase like $\lambda[\log_2 \lambda/p]/(1 + \lambda)$ it follows that the limiting efficiency is one. Thus the procedure $R_{1,D}$ for $D = 1$ is asymptotically ($N \to \infty$, $p \to 0$, $Np \to \lambda > 0$) optimal.

All the efficiencies above have been computed with respect to the ILB. If we compute efficiencies with respect to the HLB_1, we find that the procedure $R_{1,D}$ with $D = 1$ and $N = 50$ gives a result equal to the HLB in all three cases, $q = .99$, $.95$, and $.90$. Hence the procedure $R_{1,1}$ has efficiency one for these values of q.

<u>Proof of Optimality of Procedures $R_{1,D}$ and $R_{2,D}$ for $D = 1$</u>

The type of algorithm used in Section 3, equations (3.1) through (3.5) leads to an optimal NWR procedure. Hence the only question is whether a non-nested procedure can improve our results as in [3]. For $D = 1$ we have the condition (3.5) that immediately classifies the whole remaining binomial set as soon as a set is found that contains a defective unit. Hence we have only to look in the so-called defective set for the defective unit and this forces the optimal procedure to be nested. The recombination aspect can only improve our results since we only combine sets that are independent and identically distributed. Since the algorithm we use gives the best NWR procedure, the optimality result follows; this proof holds for both Models 1 and 2.

It would be desirable to have an alternate proof of the above result based on properties of the HLB; this is being investigated by a student as a thesis topic and is not yet proved. For $D \geq 2$. the HLB is not generally reached by procedure $R_{1,D}$. Cases where it is reached by procedure $R_{2,D}$ with $N = K$ are discussed in [5]; this is the case where the number of defectives (and the number of good units) is known exactly.

In Table 3 the HLB values are given for Model 1, along with the numerical values of $H_D(N)$ for $q = .75$, $.90$, $.95$, and $.99$. The efficiency relative to the HLB is easily calculated as the ratio of HLB to $H_D(N)$. For $q = .90$ this minimum (at $N = 3$) is 98.0%; for $q = .90$ this minimum (at $N = 6$) is 95.0%. For $q = .95$ this minimum (at $N = 7$) is 94.4% and for $q = .99$ this minimum (at about $N = 15$) is 94.6%. These minimum efficiencies are approximately equal for $D = 2$ and $D = 3$ and the pattern (where the minimum is attained) is also the same. Hence it appears as if the efficiency of our procedure $R_{1,D}$ under Model 1 is at least 94% for all N, all D and all q.

Although the procedure R_T (defined for $D = 1$ only) is highly efficient for $q \to 1$, the efficiency decreases as q increases and for $N = 3$ the limit $(q \to 0)$ is 54.5%.

The use of an upper bound on the number of defectives can be construed as a subtle maneuver to try to get a result that is lower than the lower bound for group-testing in the unrestricted problem. What we have done in this paper is to compare the expected number of tests (and/or the efficiencies) only for procedures that have the same probability of correctly classifying all the N units.

6. CONJECTURES AND DIRECTION OF FURTHER WORK

Under procedure $R_{1,D}$ (Model 1) for any n, any q and any $D \geq d = 1$ it is conjectured that in the minimization of (3.1) or (3.20) we can restrict our attention to two values of x, namely $x_1 = n$ for $q_0 < q \leq 1$, and

$$(6.1) \qquad x_2 = \begin{cases} n - 2^{r-1} & \text{for } 2^r \leq n \leq 3(2^{r-1}) \\ 2^r & \text{for } 3(2^{r-1}) \leq n \leq 2^{r+1} \end{cases}$$

for $0 \leq q \leq q_0$, where r is the integer part of $\log_2 n$ and q_0 is defined below. Although the minimizing value of x is generally not unique, the value of x_2 does appear in every case (of Table 1) except for the last interval $(q_0,1)$ that starts at q_0 and ends at $q = 1$, where the unique answer is to take $x = n$.

It is further conjectured that the value of the last dividing point q_0 before $q = 1$ is

$$(6.2) \qquad q_0 = \begin{cases} \dfrac{n - 2^{r-1}}{n - 2^{r-1} + 1} & \text{for } 2^r \leq n \leq 3(2^{r-1}) \\[2ex] \dfrac{2^r}{2^r + 1} & \text{for } 3(2^{r-1}) \leq n \leq 2^{r+1} \end{cases}$$

where r is the same as above.

In the interval $(q^0,1)$ where $x = n$ the value of $H_1^*(n)$ is given by

$$(6.3) \qquad H_1^*(n) = 1 + [n(r + 3) - 2^{r+1} - 1]\, p.$$

This result is not a conjecture; it follows from (3.13) and is exactly the procedure R_T mentioned above.

As pointed out above any conjectures (or results) about Model 1 lead to corresponding conjectures (or results) about Model 2.

It is hoped that in further work on this problem these conjectures will be cleared up and the optimality of procedure R_{2D} will be shown for some values of N,D and K as was shown in [5] for N=K and an infinite set of D-values.

A conjecture on the relation between Models 1 and 2 was mentioned above in Section 3 and we would like to extend that to cover other cases. For example if $D < N \leq K - D$ then we conjecture that to relate Models 1 and 2 we have to identify p^j with $D^{[j]}/(K - N + D)^{[j]}$ $(j = 0, 1, 2, \ldots)$. It would be desirable to clear up these conjectures and know how to relate these two models in the remaining cases where $K - D < N \leq D$ and where $N \geq \mathrm{Max}(D, K - D)$.

Table 1: Values of $H_D^*(N, K)$ and the Initial Strategy under Procedure $R_{1,D}$

($N = 2(1)\ 16$, $D = 1$ and $0 < q < 1$)

		q-values	Initial Strategy (or x-value)
$H_1^*(2) =$	$2 + p$	$0 < q \le 1/2$	1
	$1 + 3p$	$1/2 \le q < 1$	2
$H_1^*(3) =$	$2 + 4p$	$0 < q \le 2/3$	2
	$1 + 7p$	$2/3 \le q < 1$	3
$H_1^*(4) =$	$3 + 6p$	$0 < q \le 1/2$	2
	$2 + 8p$	$1/2 \le q \le 2/3$	2 or 3
	$1 + 11p$	$2/3 \le q < 1$	4
$H_1^*(5) =$	$3 + 10p$	$0 < q \le 1/2$	2 or 3
	$2 + 12p$	$1/2 \le q \le 3/4$	3 or 4
	$1 + 16p$	$3/4 < q < 1$	5
$H_1^*(6) =$	$3 + 14p$	$0 < q \le 1/2$	3 or 4
	$2 + 16p$	$1/2 \le q \le 4/5$	4
	$1 + 22p$	$4/5 \le q < 1$	6
$H_1^*(7) =$	$3 + 18p$	$0 < q \le 2/3$	4
	$2 + 21p$	$2/3 \le q \le 4/5$	4 or 5
	$1 + 26p$	$4/5 \le q < 1$	7
$H_1^*(8) =$	$4 + 21p$	$0 < q \le 1/2$	4
	$3 + 23p$	$1/2 \le q \le 2/3$	4 or 5
	$2 + 26p$	$2/3 \le q \le 4/5$	4 or 5 or 6
	$1 + 31p$	$4/5 \le q < 1$	8
$H_1^*(9) =$	$4 + 26p$	$0 < q \le 1/2$	4 or 5
	$3 + 28p$	$1/2 \le q \le 2/3$	4 or 5 or 6
	$2 + 31p$	$2/3 \le q \le 5/6$	5 or 6 or 7
	$1 + 37p$	$5/6 \le q < 1$	9
$H_1^*(10) =$	$4 + 31p$	$0 < q \le 1/2$	4 or 5 or 6
	$3 + 33p$	$1/2 \le q \le 2/3$	4 or 5 or 6 or 7
	$2 + 36p$	$2/3 \le q \le 6/7$	6 or 7 or 8
	$1 + 43p$	$6/7 \le q < 1$	10
$H_1^*(11) =$	$4 + 36p$	$0 < q \le 1/2$	4 or 5 or 6 or 7
	$3 + 38p$	$1/2 \le q \le 2/3$	5 or 6 or 7 or 8
	$2 + 41p$	$2/3 \le q \le 7/8$	7 or 8
	$1 + 49p$	$7/8 \le q < 1$	11
$H_1^*(12) =$	$4 + 41p$	$0 < q \le 1/2$	5 or 6 or 7 or 8
	$3 + 43p$	$1/2 \le q \le 2/3$	6 or 7 or 8
	$2 + 46p$	$2/3 \le q \le 8/9$	8
	$1 + 55p$	$8/9 \le q \le 1$	12
$H_1^*(13) =$	$4 + 46p$	$0 < q \le 1/2$	6 or 7 or 8
	$3 + 48p$	$1/2 \le q \le 3/4$	7 or 8
	$2 + 52p$	$3/4 \le q \le 8/9$	8 or 9
	$1 + 61p$	$8/9 \le q < 1$	13
$H_1^*(14) =$	$4 + 51p$	$0 < q \le 1/2$	7 or 8
	$3 + 53p$	$1/2 \le q \le 4/5$	8
	$2 + 58p$	$4/5 \le q \le 8/9$	8 or 9 or 10
	$1 + 67p$	$8/9 \le q < 1$	14
$H_1^*(15) =$	$4 + 56p$	$0 < q \le 2/3$	8
	$3 + 59p$	$2/3 \le q \le 4/5$	8 or 9
	$2 + 64p$	$4/5 \le q \le 8/9$	8 or 9 or 10 or 11
	$1 + 73p$	$8/9 \le q < 1$	15
$H_1^*(16) =$	$5 + 60p$	$0 < q \le 1/2$	8
	$4 + 62p$	$1/2 \le q \le 2/3$	8 or 9
	$3 + 65p$	$2/3 \le q \le 4/5$	8 or 9 or 10
	$2 + 70p$	$4/5 \le q \le 8/9$	8 or 9 or 10 or 11 or 12
	$1 + 79p$	$8/9 \le q \le 1$	16

Table 2: Values[§] of $H_D^*(N, K)$ and the Initial Strategy under Procedure $R_{2,D}$

$(N = 2(1)\,4,\quad D = 1 \quad \text{and} \quad N \le K < \infty)$

		K-value	Initial Strategy (or x-value)
$H_1^*(2,K) =$	$2K - 1$	$2 \le K \le 3$	1
	$K + 2$	$3 \le K < \infty$	2
$H_1^*(3,K) =$	$2K$	$3 \le K \le 5$	2
	$K + 5$	$5 \le K < \infty$	3
$H_1^*(4,K) =$	$3K - 3$	$4 \le K \le 5$	2
	$2K + 2$	$5 \le K \le 6$	2 or 3
	$K + 8$	$6 \le K < \infty$	4

[§]To get more values for this table we merely have to replace p in Table 1 by $(K - N + 1)^{-1}$, express as a single fraction, and disregard the resulting denominator. The strategies are identical for these two models and the dividing are also obtained in the same manner.

Table 3: Values of $H_D(N)$ and the HLB under Model 1 for Various N,D and q

q = .75

N	$H_1(N)$	HLB	$H_2(N)$	HLB	$H_3(N)$	HLB
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.4000	1.4000	1.6875	1.6875	1.6875	1.6875
3	1.8333	1.8333	2.4286	2.3810	2.5156	2.4688
4	2.1429	2.1429	3.0741	3.0370	3.3059	3.2471
5	2.5000	2.5000	3.6176	3.5882	4.0625	4.0000
6	2.6667	2.6667	4.1190	4.0714	4.7534	4.6849
7	2.9000	2.9000	4.5490	4.5098	5.3830	5.3138
8	3.0909	3.0909	4.9180	4.8689	5.9623	5.8787
9	3.2500	3.2500	5.2500	5.2083	6.4867	6.4067
10	3.3846	3.3846	5.5595	5.5357	6.9624	6.8898
15	3.9444	3.9444	6.7610	6.7358	8.8337	8.7500
20	4.3913	4.3913	7.6216	7.5598	10.1774	10.0876
25	4.6786	4.6786	8.2786	8.2656	11.2089	11.1312
30	4.9091	4.9091	8.8165	8.7566	12.0389	11.9924
35	5.1842	5.1842	9.2736	9.2680	12.7284	12.6446
40	5.3953	5.3953	9.6722	9.6216	13.3303	13.2887
45	5.5625	5.5626	10.0159	9.9700	13.8640	13.7744
50	5.6981	5.6981	10.3223	10.3174	14.3385	14.2960

q = .90

N	$H_1(N)$	HLB	$H_2(N)$	HLB	$H_3(N)$	HLB
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.1818	1.1818	1.2900	1.2900	1.2900	1.2900
3	1.4167	1.4167	1.6486	1.5856	1.6610	1.5980
4	1.6154	1.6154	2.0081	1.9350	2.0495	1.9685
5	1.8571	1.8571	2.3971	2.3235	2.4830	2.3955
6	2.0667	2.0667	2.7800	2.6733	2.9263	2.8051
7	2.2500	2.2500	3.1515	3.0727	3.3783	3.2875
8	2.4118	2.4118	3.5083	3.4530	3.8368	3.7466
9	2.6111	2.6111	3.8535	3.7980	4.2846	4.1779
10	2.7895	2.7895	4.1898	4.1111	4.7137	4.6221
15	3.4583	3.4583	5.4766	5.4299	6.5733	6.4850
20	3.9310	3.9310	6.5211	6.4612	8.1237	8.0167
25	4.2647	4.2647	7.3086	7.2558	9.3985	9.3073
30	4.6154	4.6154	7.9669	7.9326	10.4440	10.3661
35	4.8864	4.8864	8.5197	8.4511	11.3367	11.2415
40	5.1020	5.1020	8.9812	8.9468	12.0923	12.0140
45	5.2778	5.2778	9.3923	9.3760	12.7505	12.6744
50	5.4407	5.4407	9.7597	9.7198	13.3286	13.2135

Table 3 (cont'd)

$q = .95$

N	$H_1(N)$	HLB	$H_2(N)$	HLB	$H_3(N)$	HLB
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0952	1.0952	1.1475	1.1475	1.1475	1.1475
3	1.2273	1.2273	1.3373	1.2969	1.3401	1.2998
4	1.3478	1.3478	1.5282	1.4628	1.5375	1.4686
5	1.5000	1.5000	1.7511	1.6674	1.7705	1.6805
6	1.6400	1.6400	1.9735	1.8714	2.0074	1.8960
7	1.7692	1.7692	2.1942	2.0796	2.2480	2.1230
8	1.8889	1.8889	2.4122	2.3198	2.4920	2.3842
9	2.0357	2.0357	2.6461	2.5563	2.7551	2.6464
10	2.1724	2.1724	2.8775	2.7785	3.0200	2.9033
15	2.7353	2.7353	3.9747	3.8935	4.3630	4.2503
20	3.2564	3.2564	4.9721	4.8754	5.6833	5.5584
25	3.6818	3.6818	5.7914	5.7148	6.8226	6.6788
30	4.0204	4.0204	6.4890	6.4319	7.8760	7.7652
35	4.3519	4.3519	7.1258	7.0771	8.8447	8.7187
40	4.5763	4.5763	7.6981	7.6197	9.6967	9.5814
45	4.7656	4.7656	8.1763	8.0966	10.4698	10.3732
50	4.9565	4.9565	8.6029	8.5465	11.1773	11.0699

$q = .99$

N	$H_1(N)$	HLB	$H_2(N)$	HLB	$H_3(N)$	HLB
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0198	1.0198	1.0299	1.0299	1.0299	1.0299
3	1.0490	1.0490	1.0695	1.0599	1.0696	1.0600
4	1.0777	1.0777	1.1092	1.0907	1.1095	1.0908
5	1.1154	1.1154	1.1582	1.1311	1.1588	1.1313
6	1.1524	1.1524	1.2072	1.1717	1.2083	1.1720
7	1.1887	1.1887	1.2563	1.2126	1.2580	1.2132
8	1.2243	1.2243	1.3054	1.2564	1.3079	1.2573
9	1.2685	1.2685	1.3631	1.3078	1.3664	1.3092
10	1.3119	1.3119	1.4207	1.3595	1.4251	1.3615
15	1.5175	1.5175	1.7088	1.6208	1.7212	1.6281
20	1.7395	1.7395	2.0240	1.9270	2.0496	1.9447
25	1.9516	1.9516	2.3437	2.2291	2.3886	2.2614
30	2.1473	2.1473	2.6593	2.5413	2.7307	2.5991
35	2.3507	2.3507	2.9847	2.8749	3.0891	2.9615
40	2.5540	2.5540	3.3148	3.1976	3.4582	3.3206
45	2.7431	2.7431	3.6391	3.5104	3.8285	3.6747
50	2.9195	2.9195	3.9566	3.8204	4.1995	4.0373

Acknowledgement

The author wishes to thank Mr. Fan-Nan Lin and Mr. Jung Keun Lee both of the University of Minnesota for a number of conversations about this paper and for considerable help with the tables and calculations in the paper. Some of the conjectures in Section 6 are due to Mr. Fan-Nan Lin and will be proved separately. Thanks are also due to Ms. Elaine Frankowski for putting the Model 1 formulation on the computer, leading to the $H_D(N)$-values of Table 3 and to Mr. D. A. Florian for helping to compute the HLB-values in Table 3.

BIBLIOGRAPHY

[1] Huffman, D. A. (1952). A method for the construction of minimum redundancy codes. Proc. I. R. E. 40 1098.

[2] Sandelius, M. (1961). On an optimal search procedure. Amer. Math. Monthly 68 133 - 134.

[3] Sobel, M. and Groll, P. A. (1959). Group-testing to eliminate efficiently all defectives in a binomial sample. Bell System Tech. Journal 38 1179-1252.

[4] Sobel, M. (1960). Group-testing to classify all defectives in a binomial sample. A contribution in Information and Decision Processes. Ed. R. E. Machol. McGraw Hill, New York. (127-161).

[5] Sobel, M. (1968). Binomial and hypergeometric group-testing. Studia Scientiarum Mathematicarum Hungarica. 3 19 - 42.

[6] Thomas, J., Pasternack, B. S., Vacirca, S. J. and Thompson, D. L. (1973). Application of group-testing procedures in radiological health. (Submitted to Health Physics).